International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Cloud Mirroring: A Technique of Data Recovery

Shilpi U. Vishwakarma†* and Praveen D. Soni†

†Computer science and engineering, SGBAU, Arun colony, Amravati, Maharashtra, India

Accepted 05 March 2015, Available online 19 March 2015, **Vol.5, No.2 (April 2015)**

Abstract

Cloud computing can be termed as 'umbrella' which is used to refer as Internet based development and services. Actually Cloud computing systems represent an emerging technology that provides facilities to its users like storing their files on cloud, access it to large scale, efficient and highly reliable computing systems as pay per use. As valuable and important data of organizations are stored at a remote location on cloud we must be assured that our data is safe and be available at any time. We know that the cloud computing introduces a new type of computing platform in today's world. This type of computing will generates a large amount of private data on main cloud. Therefore, the necessity of data recovery services are growing day-by-day and it requires a development of an efficient and effective data recovery technique. The purpose of recovery technique is to help user to retrieve information from any mirror server when server lost his data and unable to provide data to the user. To achieve the recovery, many different techniques have been proposed till date. In situations like any hardware malfunction, data may get corrupted or any accidental deletion our data may no longer remain available. To maintain the data safety there must be some data recovery technique for cloud platform to recover valuable and important data efficiently in such situations mentioned above. This paper provides a new technique of data recovery by "Cloud Mirroring".

Keywords: *Cloud Computing, cloud mirroring, Recent Activity Table (RAT), message digest.*

1. Introduction

A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted virtualized, dynamically scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet. It consists of a various types of system that holds a large amount of application programs and data. It is considered as Internet-based computing where sharing and virtualizing of hardware, software and information resources are served as per on demand. Cloud computing use the internet and central remote servers to maintain data and applications and also have the ability to create, update and store files via any computer that has access to the web.

Mirroring

Cloud mirroring is nothing but the creation and maintenance of redundant copies of a cloud database. The aim is to ensure that, to provide continuous data availability and to minimize the data corruption or loss or from a situation when the operation of a network is partially compromised. Redundancy also ensures that at least one viable copy of a database will always remain accessible during system upgrades.

*Corresponding author: **Shilpi U. Vishwakarma**

2. Related Work

Cloud storage is one of the important aspect on internet, whenever user store their data they always want to retrieve their respective data from server but due to some uncertain reason the data store by the user may be get lost or corrupted but still the user want to retrieve the data. Here comes the main role of data backup and recovery. Numerous approaches have been taken by the various authors few of them are as follows,

S Sankareswari *et al*, 2014 have proposed a Seed Block Algorithm Architecture (SBA) and suggested a remote backup server. The remote Backup server is a replica of original cloud server which is physically situated at a remote location. This method is based on the concept of Exclusive-OR (XOR) operation of digital computing. The SBA uses a random number and a unique client id associated with each client.

S.S.Ganorkar *et al*, 2014 have proposed a novel data recovery service framework for cloud infrastructure, the Parity Cloud Service (PCS) provides a privacy-protected personal data recovery service. In this proposed framework user data is not required to be uploaded on to the server for data recovery. All the necessary server-side resources that provide the recovery services are within a reasonable bound.

Kruti Sharma *et al*, 2013 introduced a mechanism for online data backup technique for cloud along with

Table 1 Advantages and Disadvantages of Different Approaches

Sr. no.	Approaches	Advantages	Disadvantages
1	SBA(S Sankareswari *et al*, 2014)	Simple to implement	Inefficient
2	Parity Cloud Service(S.S.Ganorkar *et al*, 2014)	Reliable Privacy Low cost	High complexity
3	Linux Box(Kruti Sharma *et al*, 2013)	Simple Low cost	High bandwidth, Complete server backup at a time
4	HSDRT(Chi-won Song *et al*, 2011)	Used for movable clients	Costly, Increased redundancy
5	ERGOT	Exact match retrieval, privacy	Increased complexity
6	Cold/Hot Backup Strategy(Eleni Palkopoulouy *et al*, 2011)	Triggered only when failure detected	Cost increases as data increases

disaster recovery. In this approach the cost of having the backup for Cloud platform has been reduced and also it protects data from disaster at the same time the process of migration from one cloud service provider to another becomes easier and much simpler. In this approach the consumers' are not dependent on the service provider and it also eliminates the associated data recovery cost. A simple hardware box is used that achieves all these at little cost.

Chi-won Song *et al*, 2011 proposed the innovative file back-up concept HS-DRT, that makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology. This system consists of two sequences one is Backup sequence and other is Recovery sequence. The data to be backed-up is received In Backup sequence. The recovery sequence is used when there is a disaster or any data loss occurs the Supervisory starts the recovery sequence.

Vijaykumar Javaraiah *et al*, 2011 proposed Efficient Routing Grounded on Taxonomy (ERGOT) which is fully based on the semantic analysis and does not focuses on time and implementation complexity. This system is based on the Semantics that provide support for Service Discovery in cloud computing.

Yoichiro Ueno *et al*, 2010 proposed one technique that mainly focuses on the significant reduction of cost and router failure scenario i.e. (SBBR). It involves logical connectivity of IP that will be remain unchanged even after a router failure. The most important factor of this model is that it provides the network management system via multi-layer signaling.

Giuseppe Pirr´o *et al*, 2010 proposed the lowest cost point of view a model "Rent out the Rented Resources". This model is simply based on the concept of cloud vendors that rent the resources from different venture(s) and after virtualization, rents it to the clients as cloud services.

Eleni Palkopoulouy *et al*, 2011 proposed The Cold and Hot back-up strategy that performs backup and recovery on trigger basis of failure detection. In CBSRS (i.e. Cold Backup Service Replacement Strategy) recovery process, it is triggered when a service failure is detected and it will not be triggered when there is no failure. The HBSRS (i.e. Hot Backup Service Replacement Strategy), is a transcendental recovery strategy for service composition that is used for dynamic network. During the implementation of process, the backup services remains in the activated state and the first returned results of services will be used to ensure the successful implementation of service composition.

But each of these techniques has certain limitations which need to be overcome

By overviewing all of the above technique I have proposed a unique technique for data mirroring and recovery is "cloud mirroring"

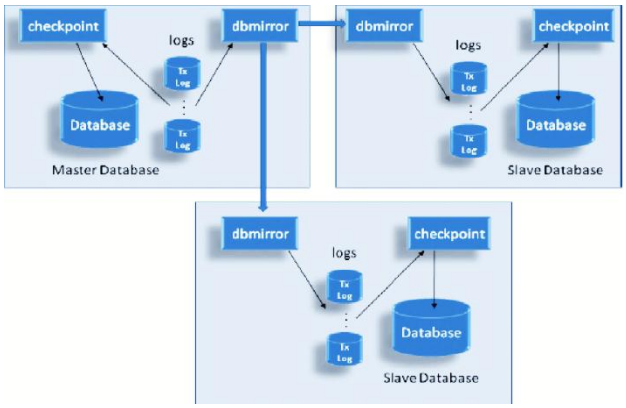

Figure 1:- Mirroring architecture (S.S.Ganorkar *et al*, 2014).

The above fig1 depicts the process of database mirroring and steps are explained as follows: - 1. Every transaction is committed to a master database by creating a log file (containing modified page images) in the database's directory. The log files are named after the transaction number they represent. Transactions are numbered serially with no gaps.

2. The checkpoint process scans the directory for log files. When one or more new log files are found, they are "checkpointed" into the database. The entire process is safe and repeatable so that there will be no loss of data.

3. To facilitate mirroring, the checkpoint process will not delete used log files, but will rename them so that they can be found by the dbmirror process.

4. The slave dbmirror process requests the "next" log from the master dbmirror process if a TFS is running, for normal transaction processing. When it receives it, it sends it to the local TFS, If a TFS is not running

together with the slave dbmirror, there may or may not be a check pointing process running. If there is, the presence of the log causes the check pointer to copy these changes into the slave database. This is repeated forever, or until the dbmirror slave process is terminated.
5. The master dbmirror process searches the database directory for log files and responds to slave dbmirrors when certain log files are requested.
6. The master dbmirror can respond to any number of slave dbmirrors.

3. Proposed Work

The proposed technique is broadly categories into two parts i.e. uploading and downloading. In the first part, user data which consist of files, documents etc. can be uploaded by the user on cloud whether it is in plain text or in encrypted format or in any other form.

Uploading module

In the uploading part, we need to maintain the log of recent activity done by the user in 'Recent Activity Table (RAT)' which stores the information of most recent file that are uploaded by the user.

Steps for uploading are as follows:

1. Firstly the user uploads his/her file on cloud; we called it as Cloud.
2. After successfully uploaded the file, file's information along with user, will store in the recent activity table (RAT).

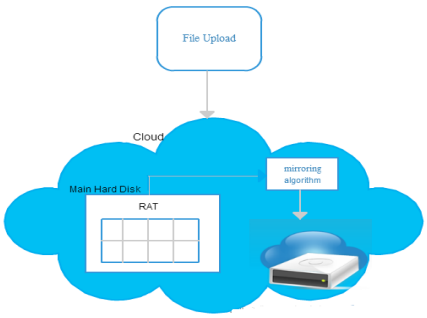

Figure 2: Uploading

- **Mirroring algorithm**

Mirroring scheduling algorithm will check the mirror copy of the user data. Mirroring starts when the CPU utilization goes below the threshold value (we assume the CPU threshold value is 50%), and daily we will do the mirroring according to time (we assume time threshold is midnight (2 a.m.)).

By using the concept of CSP (Cloud Service Provider) we will maintain the log through which we will continuously(say after 5 minutes) check the row mirror counter, after analyzing the log, CSP can dynamically change the threshold values.

Mirroring algorithm is as follows\

- **Notations**

Cpu_Threshold --- CPU Threshold
Time_Threshold --- Time Threshold
Event_Threshold --- Event Threshold
Current_Cpu --- Current Cpu
Current_Time --- Current Time
MHD --- Main Hard Disk

Pseudo code for mirroring

```
No_of_rows_mirrored= 0;
If (Current_Cpu< Cpu_Threshold)
   while(RAT.length ! empty||Current_Cpu<
                              Cpu_Threshold)
  {
Mirror the current row of RAT and delete it.
    No_of_rows_mirrored++
   }
   Return No_of_rows_mirrored;
 if(Current_Time= Time_Threshold)
while(RAT != Empty)
   {
Mirror the current row of RAT and delete it.
          No_of_rows_mirrored++;
    }
Return No_of_rows_mirrored;
```

Downloading module

The main aim of this proposed technique is to provide the recovery of user data (files) though it has been corrupted or loss etc. so the main role of mirroring technique' comes in downloading part, when user wants to download his requested data from the base cloud and if unfortunately the original data of user gets corrupted in the base cloud, then with the help of mirroring technique we will provide the same data stored by the user from mirror cloud.

Steps for downloading are as follows

1. Firstly the user will request for the user's respective data (file).
2. The user's request will moves to the retrieval algorithm where we check the presence of desired data (files), as well as its integrity.

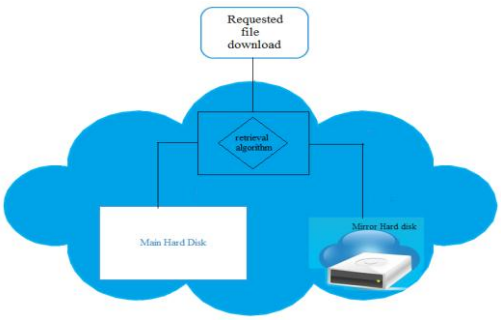

Figure 3: Downloading

File retrieval algorithm are as follows

1. When the request will arrive firstly we will check whether the requested data (files) is present in the MHD or not.
2. But if the file is corrupted and its integrity is improper then we will retrieve the respective file from mirror hard disk.

Conclusion and Future Scope

The data stored by the user is always valuable for him but no one can assure whether his data cannot be corrupted or lost so recovery plays a vital role in such scenarios. Various techniques have been proposed for data recovery but these techniques have certain limitations which need to be overcome. With the help of cloud mirroring technique we provide the high availability, integrity as well as recovery of user data (files). So for this issue we need file recovery mechanism for recovering the corrupted file. We have proposed file recovery technique by the concept of cloud mirroring. As we are using hard disk for file recovery, ultimately the cost for recovery will be reduced along with this the proposed technique is applicable to any kind of cloud. By using the cloud mirroring technique we are providing high availability to the user. This technique will focus on entire mirroring of cloud.as we are using the asynchronous mirroring the overhead of the RAT.

References

S Sankareswari, S. Hemanth (May 2014), Attribute Based Encryption with Privacy Preserving using Asymmetric Key in Cloud Computing , (IJCSIT) International Journal of Computer Science and Information Technologies.

S.S.Ganorkar, S.U.Vishwakarma, S.D.Pande (Januray 2014), An Information Security Scheme for Cloud based Environment using 3DES Encryption Algorithm, International journal of recent and development in engineering and technology(IJRDET),volume 2, issue1.

Kruti Sharma, Prof. Kavita R Singh (2013), Seed Block Algorithm: A Remote Smart Data Back-up Technique for Cloud Computing" International Conference on Communication Systems and Network Technologies IEEE.

Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang (2011), Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service, International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11.

Vijaykumar Javaraiah, Brocade Advanced Networks and Telecommunication systems (ANTS), (2011), Backupforcloud and Disaster Recovery for Consumers and SMBs, IEEE 5th International Conference.

Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki,Muzai Gakuendai, Inzai-shi, Chiba,Kazuo Ichihara, 2010, "Performance Evaluation of a Disaster Recovery System and Practical Network System Applications," Fifth International Conference on Systems and Networks Communications, pp 256-259.

Giuseppe Pirr´o, Paolo Trunfio , Domenico Talia, Paolo Missier and Carole Goble (2010), ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures, 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.

Eleni Palkopoulouy, Dominic A. Schupke, Thomas Bauscherty (2011), Recovery Time Analysis for the Shared Backup Router Resources (SBRR) Architecture, IEEE ICC.

Sheheryar Malik, Fabrice Huet (December 2011), "Virtual Cloud: Rent Out the Rented Resources," 6th International Conference on Internet Technology and Secure Transactions,11-14 ,Abu Dhabi, United Arab Emirate

Lili Sun, Jianwei An, Yang Yang, Ming Zeng (2011), Recovery Strategies for Service Composition in Dynamic Network, International Conference on Cloud and Service Computing